Published in final edited form as:

Biochim Biophys Acta. 2016 May ; 1860(5): 981–989. doi:10.1016/j.bbagen.2015.10.004.

Differential scanning calorimetry as a complementary diagnostic tool for the evaluation of biological samples

Nichola C. Garbett^{1,*} and Guy N. Brock^{2,*}

¹James Graham Brown Cancer Center, Department of Medicine, University of Louisville, Louisville, KY, USA

²Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, KY, USA

Abstract

Background—Differential scanning calorimetry (DSC) is a tool for measuring the thermal stability profiles of complex molecular interactions in biological fluids. DSC profiles (thermograms) of biofluids provide specific signatures which are being utilized as a new diagnostic approach for characterizing disease but the development of these approaches is still in its infancy.

Methods—This article evaluates several approaches for the analysis of thermograms which could increase the utility of DSC for clinical application. Thermograms were analyzed using localized thermogram features and principal components (PCs). The performance of these methods was evaluated alongside six models for the classification of a data set comprised of 300 systemic lupus erythematosus (SLE) patients and 300 control subjects obtained from the Lupus Family Registry and Repository (LFRR).

Results—Classification performance was substantially higher using the penalized algorithms relative to localized features / PCs alone. The models were grouped into two sets, the first having smoother solution vectors but lower classification accuracies than the second with seemingly noisier solution vectors.

Conclusions—Coupling thermogram technology with modern classification algorithms provides a powerful diagnostic approach for analysis of biological samples. The solution vectors

*Corresponding authors: Nichola C. Garbett, James Graham Brown Cancer Center, Department of Medicine, University of Louisville, Clinical and Translational Research Building, Room 206, 505 South Hancock Street, Louisville, KY 40202, nichola.garbett@louisville.edu, Phone: +1 502 852 3479, Fax: +1 502 852 7979. Guy N. Brock, Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Room 137, 485 East Gray Street, Louisville, KY 40202, guy.brock@louisville.edu, Phone: +1 502 852 3444, Fax: +1 502 852 3294.

DISCLOSURE

NCG is a co-inventor on pending patent applications describing the DSC plasma thermogram technology for which Louisville Bioscience, Inc. (LBI) has previously held an exclusive license from the University of Louisville. NCG is a founder, equity holder and consultant of LBI. GNB has no disclosures to declare.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

from the models may reflect important information from the thermogram profiles for discriminating between clinical groups.

General significance—DSC thermograms show sensitivity to changes in the bulk plasma proteome that correlate with clinical status. To move this technology towards clinical application the development of new approaches is needed to extract discriminatory parameters from DSC profiles for the comparison and diagnostic classification of patients.

Keywords

Diagnostic classification; Differential scanning calorimetry; Plasma proteome; Systemic lupus erythematosus; Thermogram; Penalized classification methods

1. INTRODUCTION

Differential Scanning Calorimetry (DSC) is an established biophysical technique that monitors heat capacity changes associated with the thermal denaturation of biomolecules. DSC is the only technique that directly measures the thermodynamics of intra- and intermolecular interactions stabilizing biological systems. It is a very powerful technique that has been applied to the characterization of biomolecules in a wide range of applications. A major focus has been on the rigorous determination of thermodynamic driving forces governing protein stability, folding and binding interactions as well as more qualitative thermal stability studies for drug development and biopharmaceutical formulations (e.g. [1, 2]). More recently, DSC has been applied in a new direction to the analysis of complex biological samples, such as blood plasma or other biofluids [3]. A growing number of studies suggest that differences in the thermodynamic properties of biofluid proteomes can be used to differentiate clinical samples based on health status [4–33]. It might seem initially surprising that DSC would display such sensitivity to changes in the bulk, high abundance proteome as a result of disease processes. In fact, our lab has reported that DSC changes are not related to differences in the concentration of major blood proteins in disease plasma samples. However, given the high sensitivity of DSC to modulation of biomolecule stabilization through changes in intra- and inter-molecular interactions, it is plausible that DSC biofluid profiles could reflect the modified thermal stability of major proteome components resulting from covalent modifications or binding interactions involving disease biomarkers. Although the biological mechanisms of DSC changes have not yet been reported, compelling evidence suggests the potential utility of DSC for clinical diagnostics in multiple disease settings. To realize the possibility of the clinical application of DSC, the challenge lies in demonstrating the reliable technical performance of the assay and in developing approaches to extract diagnostic information from the complex DSC profiles that can be easily applied to yield straightforward clinical metrics.

To facilitate interpretation of DSC data of clinical samples, a number of studies [9, 10, 12, 13, 16, 17, 19, 23, 24, 26] have reported the calculation of metrics that provide a readout of specific localized features of DSC profiles (e.g. heat capacity and temperature maxima of profiles). These features have been useful in discerning trends in clinical groups and calculating the statistical significance of these differences. To utilize information from the entire thermogram for diagnostic classification a number of global analysis methods have

been developed. One approach used a non-parametric method to determine differences between DSC profiles based on the distance between a test profile and averaged profiles for each class [15]. The distance was defined as the geometric average of the correlation between DSC profiles (i.e. similarity in shapes) and Euclidean distance. This approach was used to analyse DSC profiles of healthy controls and lupus patients and achieved 82% correct classification of healthy profiles and 88% for lupus. The method was developed for the analysis of plasma DSC profiles for the lupus / control data set but is generally applicable to other data sets of biofluid DSC profiles in any disease setting where any test profile can be compared to a well-defined reference group. Other groups have applied this approach for the analysis of DSC data in different settings, for example, for the classification of colorectal cancer based on DSC profiles [24]. Another method for the analysis of plasma profiles employed a parametric statistical model developed for the classification of cervical cancer versus healthy controls [22]. Here, DSC profiles were reduced in complexity by restricting the temperature range to that encompassing the major heat capacity signal (50–76°C) and averaging over 1°C temperature increments. Profiles were then subjected to a logarithmic transformation and fit to a linear regression model. This method performed extremely well for the healthy / cervical cancer data set with a mean classification rate of 97%. As the model used in this approach was developed from this specific data set, the utility of this method requires the evaluation of this approach with other data sets. Another useful method was based on deconvoluting the DSC profile into several component curves, each with a defined height, center and width which were used in a multiparametric analysis for the classification of healthy controls and gastric adenocarcinoma patients [27]. The construction of polygonal plots from these three parameters for each of the component curves provided a useful graphical tool to distinguish patient groups. Also, similar to some earlier reports, the area and first moment, or average, temperature of DSC profiles were found to display differences between the controls and gastric adenocarcinoma patients.

The approaches discussed above demonstrate the evolution in the development of analytical tools for the characterization of DSC biofluid profile features associated with various clinical conditions. More work is needed to develop and validate reliable analytical approaches to provide a rapid, easily interpretable diagnostic result that can be readily employed in the clinical setting. The focus of this report is the application of new analysis approaches to DSC data for the purpose of diagnostic classification.

2. MATERIALS AND METHODS

2.1 Plasma samples

De-identified plasma samples and patient data were obtained from the Lupus Family Registry and Repository (LFRR) [34]. Plasma samples for 300 patients meeting the revised criteria of the American College of Rheumatology for SLE [35] and 300 healthy controls matched demographically by sex, ethnicity and age were received and kept at –80 °C until thawed for DSC analysis.

2.2 Collection of DSC thermograms

DSC samples were prepared and analyzed according to our previously published procedure [10] which includes a detailed account of our experimental procedures. Data were collected using an automated VP-Capillary DSC system (MicroCal, LLC, Northampton, MA, now a division of Malvern Instruments Inc.). Electrical calibration of the differential power signal and temperature calibration using hydrocarbon temperature standards were performed as part of the manufacturer periodic instrument maintenance. Interim instrument performance was assessed using biological standards lysozyme and RNaseA. Samples and dialysate were loaded into 96 well plates thermostated at 5 °C within the instrument autosampler until analysis. Thermograms were recorded from 20 °C to 110 °C at a scan rate of 1 °C/min with a pre-scan thermostat of 15 minutes, mid feedback mode and a filtering period of 2 seconds. Duplicate thermograms were obtained for each plasma sample. DSC data were analyzed using Origin 7 (OriginLab Corporation, Northampton, MA). Raw DSC data were corrected for the instrumental baseline by subtraction of a suitable buffer reference scan. Thermograms were normalized for the total protein concentration and corrected for non-zero baselines by application of a linear baseline fit. Final thermograms were plotted as excess specific heat capacity (J/K.g) versus temperature (°C).

2.3 Summary metrics of DSC thermograms

Thermograms are frequently characterized by metrics summarizing the shape and prominent features of the thermograms [9]. These include: (1) the total area under the thermogram (typically from 45–90°C); (2) the maximum excess specific heat capacity at various peaks (e.g. Peak 1 height, Peak 2 height, etc.); (3) the overall maximum peak height; (4) the width of the primary thermogram peak at half height; (5) the temperature of the peak maximum ($T_{\max}$); (6) the ratio of the peak heights (e.g., (Peak 1 height) / (Peak 2 height), etc.); and (7) the “mean” or first moment temperature of the thermogram, T_{FM} , where

$$T_{FM} = \frac{\int_{45}^{90} (TC_p^{ex}) dT}{\int_{45}^{90} C_p^{ex} dT}$$

and C_p^{ex} represents the excess specific heat capacity at a given temperature. These summary metrics can be used in lieu of the original thermogram values for classifying disease status based on any of the classification models described below. While the calculated metrics are useful for characterizing certain aspects of the thermograms, they are not necessarily informative for differences between patient classes.

Principal components (PCs) are another common technique for summarizing the information in a data matrix in a concise manner. PCs are not intended as a classification technique per se, but are commonly used as a dimension reducing tool prior to building a classification or regression model. PCs are the set of orthogonal vectors or factors such that the first vector is the direction which explains the most variation in the data, the second vector is the direction which explains the second greatest percentage of variation in the data that is orthogonal to the first, and so on. The solution can be obtained from the eigenvalue decomposition of the

covariance matrix of the data, where the principal components directions are the eigenvectors of the covariance matrix and the variance of the principal components are proportional to the eigenvalues of the covariance matrix. Typically, only the number of components needed to explain 90–95% of the total variation in the data are retained (e.g., as determined by the “elbow” in a scree plot). Once the principal components are determined, these can be used as a replacement for the original variables in a classification problem. The PCs have the advantage that they are orthogonal and hence avoid computational issues associated with multicollinearity. However, they are not specifically designed for a classification problem and do not make explicit use of the clinical classification of the data in their construction. Hence they are frequently sub-optimal for classification problems.

2.4 Classification methods

The thermogram values at each temperature can be treated as variables and used to develop classification models for diseased versus healthy individuals. Since the number of thermogram values is large and can potentially lead to overfitting, variable selection techniques should be employed to reduce the dimension of the problem. This can be accomplished via sparse or penalized methods, which are typically available for the more commonly used classification methods. The goal is to use the information in the thermogram profiles to classify a patient as having a disease (here, lupus) or not. That is, to develop a phenomenological model governed by a set of parameters that can be used to predict the class label (i.e. disease or not) of each thermogram. The unconstrained solutions to the problem (i.e., the parameter estimates associated with each of the predictor variables) are typically obtained by minimizing a suitable objective function (e.g., for statistical models this is typically the negative log-likelihood function). The idea behind penalized methods is to employ a penalty in the objective function which prevents overly-complex solutions. These penalty functions are based on the magnitude of the coefficient vector, and can either shrink the coefficients overall (the ridge or L_2 penalty) or eliminate some of the coefficients entirely (the ‘least absolute shrinkage and selection operator’ (lasso) or L_1 penalty). The latter penalty function is a form of variable selection since some of the coefficients are shrunk to zero (and thus eliminated from the model). Other possibilities for penalty functions exist, including the elastic net which is a compromise combining the ridge and lasso penalty functions and tends to retain or discard groups of correlated variables together. The degree of penalization is controlled by a parameter which varies from very stringent (e.g., all parameters are shrunk to zero) to nonexistent (the unconstrained solution). The optimal level of shrinkage or penalization is usually determined empirically by a cross-validation process. A good introduction to penalized methods for classification is available in Hastie, Tibshirani, and Friedman, especially Chapter 18 [36]. Below is a brief description of the classification methods and software packages that were used for analyzing the DSC thermograms in this study.

2.4.1 Logistic regression (LR)—Logistic regression is an example of a generalized linear model (GLM), which extends the statistical theory for linear models to the case where the response variable (here, heat capacity) is non-normally distributed. In this context the logistic regression model models the probability that an individual will have lupus, given the set of input thermogram values at specific temperatures. Predicted probabilities can be

obtained from the model and used to classify patients as having lupus or not (e.g., using a threshold probability of 0.5). Penalized solutions to the problem are available in a number of R packages including *lqa* [37] and *glmnet* [38]. The *glmnet* package uses the elastic net penalty and allows users to select between the lasso, ridge, or any weighted combination of the two penalties. In this work we use the *glmnet* package with both the lasso (LR-LASSO) and elastic net (LR-ENET, equally weighted combination of lasso and ridge) penalties.

2.4.2 Support vector machines (SVM)—Support vector machines have enjoyed great success in classification problems since their introduction [39]. The idea behind SVMs is to find the hyperplane in multidimensional space such that the margin (or separation) between the training points for the two classes is maximized. While SVMs do not return a predicted probability, historically they have been very successful for classification problems. Several potential advantages of SVMs include the focus on points (subjects) which characterize the boundary between two classes and the ability to incorporate features mapped into a different space via a kernel function (a function which computes the similarity or proximity of two points in the transformed space). In this work we use the *penalizedSVM* package in R [40] to fit penalized SVM models using a linear kernel. In addition to the lasso penalty, the package also implements the Smoothly Clipped Absolute Deviation (SCAD) penalty. The SCAD penalty behaves similarly to the lasso for small coefficients but retains the large coefficients as they are. We evaluate both the SCAD penalized SVM model (SVM-SCAD) and the SCAD penalty combined with an L_2 penalty [41], the elastic SCAD (SVM-ESCAD).

2.4.3 Fisher's linear discriminant analysis (LDA)—The goal of Fisher's LDA is to find the direction in the covariate space that best separates the two (or more) classes of patients. That is, the linear combination of the covariates which maximizes the ratio of the between group variation to the within group variation. When the number of covariates (here, heat capacity values at each temperature) far exceeds the number of subjects, classical LDA cannot be directly applied and a regularized or penalized approach is needed. This can be accomplished by shrinking the covariance matrix to the more stable and easily invertible identity matrix (a matrix consisting of ones on the diagonal and zeros elsewhere) and substituting this into the LDA algorithm [42]. Other authors included an L_1 penalty or L_1 constraint on the objective function for LDA to enforce dimension reduction (sparsity) on the coefficients of the resulting discriminant vector [43, 44]. R packages which employ the latter approaches include *penalizedLDA* [45] and *MGSDA* [46]. In this work we evaluate the L_1 or lasso constrained LDA as estimated by the *MGSDA* package (LDA-LASSO).

2.4.4 Partial least squares (PLS)—Partial least squares is similar to principal components, but instead of finding orthogonal factors that explain the most variation in the covariates the components are determined which find the greatest covariance between the covariates and the response variable. In this case, the first PLS component is the linear combination of the thermogram values which has the strongest covariance with the disease status (lupus or normal), and subsequent components orthogonal to the first are determined in an analogous fashion. PLS has a long history of application in chemometrics (see [47] for a recent review of application within metabolomics data). Similarly to the above classifiers, penalized versions achieving a sparse coefficient or loading vectors are obtained by directly

penalizing the objective function. When the response variable is binary, a version of PLS called PLS-DA (for PLS discriminant analysis) is applied. In the R package *spls* this is achieved by a two-step process where sparse PLS is first applied for dimension reduction (sparsity) and PLS dimensions are subsequently used in an LDA or logistic regression classifier [48, 49]. In this work we evaluate the sparse PLS-DA method as obtained by the *splsda* function in package *spls* (SPLS-DA).

3. RESULTS

To develop preliminary approaches for enhanced data analyses of DSC data we utilized our most substantial DSC data set collected using samples from the Lupus Family Registry and Repository (LFRR) which is comprised of 300 lupus patients and 300 demographically-matched controls. We examined an array of approaches to characterize thermogram differences related to clinical status in this large data set and determined the performance of each approach for the diagnostic classification of patients based on DSC data. Two case samples and six control samples were flagged as poor quality data and removed prior to analysis.

Figure 1 (left panel) displays the median thermogram profiles for both lupus and control patients along with empirical 10th and 90th percentiles at each temperature. A casual inspection reveals prominent differences between the two profiles at the first peak (62–67°C) and more subtle differences at a third peak around 75–80°C. The second peak (69–73°C) is more similar between the two groups, though still statistically significantly different ($p < 0.001$, t-test). These visual findings are corroborated by the first principal component (PC) (Figure 1, right panel), which contrasts differences between the heights of the first and third peaks. That is, subjects with larger values of peak 3 and smaller values of peak 1 (e.g., lupus patients) will also have larger values of the first PC. The second PC has all positive loadings and is associated with the total area under the thermogram curve. The third and fourth PCs are more difficult to interpret, but seemingly involve contrasting thermogram values primarily in the 60–62°C, 69–71°C, 74–76°C, and 80–82°C ranges.

Scatter plot matrices for the maximum peak heights of peaks 1, 2, and 3 are shown in Figure 2, color-coded by lupus or control status. The figure reiterates what was observed in Figure 1, namely that peaks 1 and 3 differ between lupus patients and controls while peak 2 largely does not. However, there is considerable overlap between the two groups of subjects in the scatter plots. Figure 3 similarly plots scatter plot matrices for $T_{\max}$, T_{FM} , and the (peak 1) / (peak 3) ratio. While there are interesting distributional patterns, only the (peak 1) / (peak 3) ratio offers any substantial separation between the groups. However, all the summary metrics were statistically significantly different ($p < 0.05$, t-test) between lupus cases and controls with the exception of the width at half height.

The utility of the summary metrics and PCs for patient classification can be further illustrated via receiver operating characteristic (ROC) curves. For each possible cut-point of the summary metrics / PCs, patients are classified as having lupus or not (e.g., based on whether they are above / below the cut-point). The sensitivity and specificity of the resulting classification is determined by comparing to the true disease status. ROC curves are then

constructed by plotting the resulting sensitivity and one minus the specificity values for all of these cut-points. The area under the curve (AUC) gives an overall measure of predictive ability called the *concordance* index or C-index, with values of 0.5 corresponding to a random guess and values of 1.0 corresponding to perfect separation. Figure 4 plots the ROC curves for the first six PCs, while Figure 5 plots ROC curves for six of the calculated summary metrics. Figure 4 shows that while the first PC is useful for separating patients (AUC = 0.78), the remaining PCs are not discriminatory with a maximum AUC value of 0.62. Likewise, Figure 5 shows that peak 1 and the (peak 1)/(peak 3) ratio are useful discriminators (AUC $\approx$ 0.80) while peak 3, $T_{\max}$, and T_{FM} are useful to a lesser extent (AUC between 0.71 and 0.74) and peak 2 is not particularly predictive (AUC = 0.59).

To evaluate the classification accuracy of models based on PCs and summary metrics we divided the subjects into a training set (2/3rds) and a test set (remaining 1/3rd). A logistic regression model was fitted using the first six PCs based on the training data, and likewise a similar model was fitted using five of the six summary metrics displayed in Figure 5 (peak 1, peak 3, (peak 1)/(peak 3) ratio, $T_{\max}$, and T_{FM}). The test data accuracy for the two models was very similar: 70.9% for the PC model and 70.4% for the summary metrics model.

We subsequently evaluated the performance of the six classification models: LR-LASSO, LR-ENET, SVM-SCAD, SVM-ESCAD, LDA-LASSO, and SPLS-DA. To capture the most informative range of the thermogram the temperature was restricted to between 60.0 and 80.9°C. Since thermogram values were recorded in 0.1°C increments, this resulted in 210 total features available for classification purposes. To assess variability in the classification accuracy and the solution vector of the methods we randomly split the data 100 times into a 2/3rds training set and 1/3rd test set. The overall accuracy for the six models on the 100 test data sets are displayed in Figure 6. The best performing models were clearly LR-ENET (median accuracy 88%), LR-LASSO (87%), and LDA-LASSO (87%). This was followed by SVM-SCAD (83%), while SVM-ESCAD (74%) and SPLS-DA (74%) had the lowest test data accuracy levels. Note however that all of the models outperformed classification based on the PCs and summary metrics.

Since only two classes of subjects were being compared, in each case the solution for the classifier resulted in a single discriminatory variable calculated as a weighted average of the thermogram values. The weights correspond to the solution vector (e.g., the coefficient vector) and are plotted in Figure 7 across the 100 splits of the data. Note that the dimension reduction (sparsity) criterion coincides with the zero coefficients in the solution, which is particularly evident in the lasso constrained solutions. Also, there is a remarkable degree of similarity in the solution pattern for LR-LASSO, LR-ENET, LDA-LASSO, and to a lesser extent SVM-SCAD. In particular, the pattern between 67 and 72.5°C is consistently maintained across all the methods (the green shaded region in each of the plots gives the 10th and 90th percentiles for the solution vector across the 100 data splits). Patterns around 60°C, 65°C, 77°C, and 80°C are also fairly well maintained. In contrast, the solution patterns for SVM-ESCAD and SPLS-DA are decidedly different from the other four classifiers and also similar to each other. While the solution for these two classifiers appears 'smoother' compared to the other four, the classification accuracy is notably lower (Figure 6). A final note concerns the difference in magnitude of the coefficients between SVM-

ESCAD / SPLS-DA and the other four classifiers. However, the large magnitude of the coefficients for these four classifiers does not result in high variability in classification accuracy, as evidenced by Figure 6.

4. DISCUSSION

In this paper, we describe the use of DSC thermograms as a diagnostic tool and illustrate its application for classifying lupus cases versus controls. We compared classification accuracy based on summary metrics of the thermograms with classification algorithms specifically tuned to distinguish lupus cases from controls based on the thermogram information. Penalized methods were used to constrain the solution and reduce the dimension of the problem. Our results indicate that substantially improved performance is obtained with the classification algorithms relative to summary metrics / PCs alone, particularly for LR-LASSO, LR-ENET, and LDA-LASSO.

The strongest performing model (LR-ENET) has a median accuracy of 88% based on left-out test samples with similar sensitivity (88%) and specificity (88%). This balanced sensitivity and specificity is not surprising given the balanced nature of the sample, and the diagnostic value of thermograms will ultimately be determined by their actual performance in the clinic. However, these results are at least competitive with antinuclear antibody (ANA) testing for systemic lupus, which is highly specific (97%) but has low sensitivity (57%) (c.f. Table 2 in [50]). A further note is that the classification performance of the thermograms was achieved via comparison with a heterogeneous mixture of control subjects with a variety of co-morbidities including rheumatoid arthritis, osteoarthritis, anemia, high blood pressure, and diabetes, among others. This heterogeneous mixture of control subjects may also partially explain why the relative peak two / peak one height among control subjects was higher than what we have observed in other studies. Our current work is focusing on coupling thermogram data with ANA testing to improve diagnostic classification of lupus relative to ANA testing alone.

In contrasting the results from the different classification algorithms, the solutions could be grouped into two sets of high similarity. While the coefficient vectors for SVM-ESCAD and SPLS-DA were fairly smooth and seemingly easier to interpret, the classification accuracies based on the LDA, SVM-SCAD, and LR methods were substantially better. Though at first glance the coefficient vectors obtained for LDA-LASSO, LR-ENET, LR-LASSO, and SVM-SCAD appear 'noisy', the solutions were very similar to each other and relatively consistent across multiple splits of the data. Thus these coefficient patterns may relay important information concerning contrasting elements of the thermogram profiles that distinguish diseased (here, lupus) and healthy individuals. Theoretical connections between the loss functions (objective functions) for LDA, LR, and SVM are discussed in Section 12.3.2 in [36].

To handle the dimensionality of the problem, we used penalized methods to simultaneously obtain the solution vector and select important thermogram values for classification. Other options for variable selection include filter and wrapper methods. Filter methods involve selecting variables based on a univariate test statistic (e.g., a t-test or Wilcoxon test for

Author Manuscript

differences between cases and controls) applied to each variable. The variables with the most significant results (usually based on a pre-defined p-value threshold) are then used for classification. While simple to apply, the combination of variables selected are not necessarily ideal for classification. In contrast, wrapper approaches are designed to select variables optimal for a particular classification algorithm. This is accomplished by defining subsets consisting of a decreasing number of variables, where for each subset the variables are ranked by a variable importance measure and the least significant predictors are removed to obtain the next subset. In order to avoid over-fitting, such approaches are typically wrapped within a double cross-validation scheme. Wrapper methods are thus more comparable to penalized approaches, but may be more computationally burdensome (depending on the coarseness of the number of subsets evaluated). The *caret* package in R [51] is a good resource for applying many classification algorithms coupled with wrapper selection and variable importance measures.

Author Manuscript

The thermogram value at each 0.1°C temperature increment were used as input to allow maximum flexibility for the classifiers to select which thermogram values were most informative for segregating subjects. This represents an important initial step in determining what regions of the thermogram differ critically between cases and controls, and how these regions ‘interact’ or ‘contrast’ (i.e., as indicated by their coefficients). However, interpretation of the resulting coefficient profiles was challenging in certain cases, e.g. for the lasso penalized solutions. Hence, further work on decomposing the thermograms into salient and constituent peaks prior to applying classification approaches may improve model understanding while retaining full diagnostic utility [27]. Such approaches are planned as future research.

Author Manuscript

There are several general comments concerning diagnostic classification models that need to be reiterated here. First, statistical significance does not imply clinical relevance or importance for predictive accuracy. With sufficient sample size, even minor differences that have low discriminatory power will appear statistically significant. This issue is exemplified by the differences in the heights of peak two between lupus subjects and controls. This difference is statistically significant due to the large sample size in our study, but has relatively limited predictive ability (C-index of 0.59). Hence, when evaluating the utility of a diagnostic tool, predictive ability in addition to statistical significance *must* be considered. Second, classification accuracy must always be evaluated using an independent test set. Classification methods are particularly adept at finding solutions to discriminate between class labels, and evaluating the models based on data re-substitution will result in overly-optimistic (and often misleading) conclusions.

Author Manuscript

The results from this study show the critical importance in the development of diagnostic methods for the classification of clinical thermogram data. The growing number of studies applying DSC in multiple disease settings has served to illustrate the potential utility of DSC in characterizing clinical samples. Initial studies focused on straightforward approaches to correlate changes in thermogram features with clinical groups [9, 10, 12, 13, 16, 17, 19, 23, 24, 26]. Although consideration of certain thermogram features is useful in examining differences between groups, this study has shown that classification performance based on such measures (e.g., summary metrics) is limited. This study has evaluated a number of

approaches for the diagnostic classification of DSC data but further development is needed in translating DSC towards clinical application. These approaches would also have to be amendable for clinical implementation in terms of generating a readily interpretable diagnostic result appropriate for the clinic setting. It is also critical to discover the association between biological disease processes and thermogram changes. Our prior studies have identified the “assignment” of peaks in the healthy thermogram through the study of individual purified plasma proteins [5]. The situation in the disease state is much more complicated where modified thermal stabilities of major plasma proteins resulting from biomarker processes would result in complex thermogram changes. The accurate representation of these changes would serve to deconvolute the thermogram disease signature and provide an enhanced diagnostic approach focused on particular components or regions of the thermogram.

In conclusion, we have demonstrated that thermogram technology coupled with modern classification algorithms provides a powerful diagnostic approach for analysis of biological samples. Future work remains to develop an algorithm that is simultaneously interpretable while maintaining a high performance level. Uncovering the biological phenomena that drive the thermogram changes associated with a disease state will also lead to enhanced diagnostic approaches as well as make important biological discoveries which could improve our understanding of the underlying disease etiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge the immensely valuable resource of the Lupus Family Registry and Repository for generously providing access to patient samples and data for this study. We thank the James Graham Brown Cancer Center for support of the plasma thermogram research. Collection of plasma thermogram data was supported by use of the Biophysical Core Facility at the James Graham Brown Cancer Center and by grants P20RR018733 and P20GM103482 from the National Institutes of Health. Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R21CA187345 (NCG) and by a Competitive Enhancement Grant from the Office of the Executive Vice President for Research and Innovation at the University of Louisville (NCG).

ABBREVIATIONS

AUC	area under the curve
C_p^{ex}	excess specific heat capacity
DSC	differential scanning calorimetry
GLM	generalized linear model
LASSO	least absolute shrinkage and selection operator
LDA	linear discriminant analysis
LDA-LASSO	lasso constrained LDA model
LFRR	Lupus Family Registry and Repository

LR	logistic regression
LR-LASSO	lasso constrained LR model
LR-ENET	elastic net constrained LR model
PC	principal component
PLS	partial least squares
PLS-DA	PLS discriminant analysis
ROC	receiver operating characteristic
SCAD	smoothly clipped absolute deviation
SLE	systemic lupus erythematosus
SPLS-DA	sparse PLS-DA
SVM	support vector machines
SVM-ESCAD	elastic SCAD penalized SVM model
SVM-SCAD	SCAD penalized SVM model
T_{FM}	first moment temperature
T_{max}	temperature of the peak maximum

References

- Cooper, A.; Nutley, MA.; Wadood, A. Differential scanning microcalorimetry. In: Chowdhry, BZ.; Harding, SE., editors. Protein-Ligand Interactions: hydrodynamics and calorimetry: a practical approach. Oxford University Press; Oxford, UK: 2001. p. 287-318.
- Johnson CM. Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys. 2013; 531:100–109. [PubMed: 23022410]
- Garbett NC, Mekmaysy CS, DeLeeuw L, Chaires JB. Clinical application of plasma thermograms. Utility, practical approaches and considerations. Methods. 2015; 76:41–50. [PubMed: 25448297]
- Garbett NC, Mekmaysy CS, Helm CW, Jenson AB, Chaires JB. Differential scanning calorimetry of blood plasma for clinical diagnosis and monitoring. Exp Mol Pathol. 2009; 86:186–191. [PubMed: 19146849]
- Garbett NC, Miller JJ, Jenson AB, Chaires JB. Calorimetry outside the box: a new window into the plasma proteome. Biophys J. 2008; 94:1377–1383. [PubMed: 17951300]
- Garbett NC, Miller JJ, Jenson AB, Chaires JB. Calorimetric Analysis of the Plasma Proteome. Semin Nephrol. 2007; 27:621–626. [PubMed: 18061844]
- Garbett NC, Miller JJ, Jenson AB, Miller DM, Chaires JB. Interrogation of The Plasma Proteome with Differential Scanning Calorimetry. Clin Chem. 2007; 53:2012–2014. [PubMed: 18030697]
- Garbett NC, Miller JJ, Jenson AB, Chaires JB. Ligand Binding Alters the Calorimetric Thermogram of Albumin. J Clin Ligand Assay. 2006; 29:194–197.
- Garbett NC, Merchant ML, Chaires JB, Klein JB. Calorimetric analysis of the plasma proteome: Identification of type 1 diabetes patients with early renal function decline. Biochim Biophys Acta. 2013; 1830:4675–4680. [PubMed: 23665587]
- Garbett NC, Merchant ML, Helm CW, Jenson AB, Klein JB, Chaires JB. Detection of Cervical Cancer Biomarker Patterns in Blood Plasma and Urine by Differential Scanning Calorimetry and Mass Spectrometry. PLOS ONE. 2014; 9:e84710. [PubMed: 24416269]

11. Chagovetz AA, Jensen RL, Recht L, Glantz M, Chagovetz AM. Preliminary use of differential scanning calorimetry of cerebrospinal fluid for the diagnosis of glioblastoma multiforme. *J Neurooncol.* 2011; 105:499–506. [PubMed: 21720810]
12. Chagovetz AA, Quinn C, Damarse N, Hansen LD, Chagovetz AM, Jensen RL. Differential scanning calorimetry of gliomas: a new tool in brain cancer diagnostics? *Neurosurgery.* 2013; 73:289–295. [PubMed: 23624408]
13. Fekecs T, Zapf I, Ferencz A, L rinczy D. Differential scanning calorimetry (DSC) analysis of human plasma in melanoma patients with or without regional lymph node metastases. *J Therm Anal Calorim.* 2012; 108:149–152.
14. Ferencz, A.; Fekecs, T.; L rinczy, D. Differential Scanning Calorimetry, as a New Method to Monitor Human Plasma in Melanoma Patients with Regional Lymph Node or Distal Metastases. In: Xi, Y., editor. *Skin Cancer Overview*. InTech; 2011. p. 141-151.
15. Fish DJ, Brewood GP, Kim JS, Garbett NC, Chaires JB, Benight AS. Statistical analysis of plasma thermograms measured by differential scanning calorimetry. *Biophys Chem.* 2010; 152:184–190. [PubMed: 20961680]
16. Krumova S, Rukova B, Todinova S, Gartcheva L, Milanova V, Toncheva D, Taneva SG. Calorimetric monitoring of the serum proteome in schizophrenia patients. *Thermochim Acta.* 2013; 572:59–64.
17. Mehdi M, Fekecs T, Zapf I, Ferencz A, L rinczy D. Differential scanning calorimetry (DSC) analysis of human plasma in different psoriasis stages. *J Therm Anal Calorim.* 2013; 111:1801–1804.
18. Michnik, A. Blood plasma, serum and serum proteins microcalorimetric studies aimed at diagnosis support. In: L rinczy, D., editor. *Thermal Analysis in Medical Application*. Akadémiai Kiadó; Budapest, Hungary: 2011. p. 171-190.
19. Michnik A, Drzazga Z, Michalik K, Barczyk A, Santura I, Soza ska E, Pierzchała W. Differential scanning calorimetry study of blood serum in chronic obstructive pulmonary disease. *J Therm Anal Calorim.* 2010; 102:57–60.
20. Michnik A, Drzazga Z, Poprzecki S, Czuba M, Kempa K, Sadowska-Krepa E. DSC serum profiles of sportsmen. *J Therm Anal Calorim.* 2013; 113:365–370.
21. Moezzi M, Ferencz A, L rinczy D. Evaluation of blood plasma changes by differential scanning calorimetry in psoriatic patients treated with drugs. *J Therm Anal Calorim.* 2014; 116:557–562.
22. Rai SN, Pan J, Cambon A, Chaires JB, Garbett NC. Group Classification based on High-Dimensional Data: Application to Differential Scanning Calorimetry Plasma Thermogram Analysis of Cervical Cancer and Control Samples. *Open Access Medical Statistics.* 2013; 3:1–9.
23. Todinova S, Krumova S, Gartcheva L, Robeerst C, Taneva SG. Microcalorimetry of blood serum proteome: a modified interaction network in the multiple myeloma case. *Anal Chem.* 2011; 83:7992–7998. [PubMed: 21928840]
24. Todinova S, Krumova S, Kurtev P, Dimitrov V, Djongov L, Dudunkov Z, Taneva SG. Calorimetry-based profiling of blood plasma from colorectal cancer patients. *Biochim Biophys Acta.* 2012; 1820:1879–1885. [PubMed: 22903026]
25. Wisniewski M, Garbett NC, Fish DJ, Brewood GP, Miller JJ, Chaires JB, Benight AS. Differential Scanning Calorimetry in Molecular Diagnostics. *In Vitro Diagnostic Technology.* 2011; 17:29–34.
26. Zapf I, Fekecs T, Ferencz A, Tizedes G, Pavlovics G, Kálmán E, L rinczy D. DSC analysis of human plasma in breast cancer patients. *Thermochim Acta.* 2011; 524:88–91.
27. Vega S, Garcia-Gonzalez MA, Lanás A, Velazquez-Campoy A, Abian O. Deconvolution Analysis for Classifying Gastric Adenocarcinoma Patients Based on Differential Scanning Calorimetry Serum Thermograms. *Sci Rep.* 2015; 5:7988. [PubMed: 25614381]
28. Kikalishvili L, Ramishvili M, Nemsadze G, Lezhava T, Khorava P, Gorgoshidze M, Kiladze M, Monaselidze J. Thermal stability of blood plasma proteins of breast cancer patients, DSC study. *J Therm Anal Calorim.* 2015; 120:501–505.
29. Zapf I, Moezzi M, Fekecs T, Nedvig K, L rinczy D, Ferencz A. Influence of oxidative injury and monitoring of blood plasma by DSC on breast cancer patients. *J Therm Anal Calorim.* 201510.1007/s10973-10015-14642-10979

30. Krumova S, Todinova S, Danailova A, Petkova V, Dimitrova K, Gartcheva L, Taneva SG. Calorimetric features of IgM gammopathies. Implication for patient's diagnosis and monitoring. *Thermochim Acta*. 2015; 615:23–29.
31. Barceló F, Cerdà JJ, Gutiérrez A, Jimenez-Marco T, Durán MA, Novo A, Ros T, Sampol A, Portugal J. Characterization of Monoclonal Gammopathy of Undetermined Significance by Calorimetric Analysis of Blood Serum Proteome. *PLOS ONE*. 2015; 10:e0120316. [PubMed: 25794164]
32. Moezzi M, Zapf I, Fekecs T, Nedvig K, Linczy D, Ferencz A. Influence of oxidative injury and monitoring of blood plasma by DSC on patients with psoriasis. *J Therm Anal Calorim*. 2015; 100:1007/s10973-10015-14674-10971
33. Szalai Z, Molnár TF, Linczy D. Differential scanning calorimetry (DSC) of blood serum in chronic obstructive pulmonary disease (COPD). *J Therm Anal Calorim*. 2013; 113:259–264.
34. Rasmussen A, Sevier S, Kelly JA, Glenn SB, Aberle T, Cooney CM, Grether A, James E, Ning J, Tesiram J, Morrissey J, Powe T, Drexel M, Daniel W, Namjou B, Ojwang JO, Nguyen KL, Cavett JW, Te JL, James JA, Scofield RH, Moser K, Gilkeson GS, Kamen DL, Carson CW, Quintero-del-Rio AI, del Carmen Ballesteros M, Punaro MG, Karp DR, Wallace DJ, Weisman M, Merrill JT, Rivera R, Petri MA, Albert DA, Espinoza LR, Utset TO, Shaver TS, Arthur E, Anaya JM, Bruner GR, Harley JB. The lupus family registry and repository. *Rheumatology*. 2011; 50:47–59. [PubMed: 20864496]
35. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum*. 1997; 40:1725. [PubMed: 9324032]
36. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. 2. Springer; New York: 2009.
37. Ulbricht, J. R package version 1.0-3. 2012. lqa: Penalized Likelihood Inference for GLMs.
38. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010; 33:1–22. [PubMed: 20808728]
39. Vapnik, V. *The Nature of Statistical Learning Theory*. Springer; New York: 1996.
40. Becker N, Werft W, Toedt G, Lichter P, Benner A. penalizedSVM: a R-package for feature selection SVM classification. *Bioinformatics*. 2009; 25:1711–1712. [PubMed: 19398451]
41. Becker N, Toedt G, Lichter P, Benner A. Elastic SCAD as a novel penalization method for SVM classification tasks in high-dimensional data. *BMC Bioinformatics*. 2011; 12:138. [PubMed: 21554689]
42. Xu P, Brock GN, Parrish RS. Modified linear discriminant analysis approaches for classification of high-dimensional microarray data. *Comput Stat Data Anal*. 2009; 53:1674–1687.
43. Witten DM, Tibshirani R. Penalized classification using Fisher's linear discriminant. *J R Stat Soc Series B Stat Methodol*. 2011; 73:753–772. [PubMed: 22323898]
44. Gaynanova I, Booth JG, Wells MT. Simultaneous sparse estimation of canonical vectors in the $p \gg N$ setting. *J Am Stat Assoc*. 2015; 110:1080/01621459.01622015.01034318
45. Witten, D. R package version 1.0. 2011. penalizedLDA: Penalized classification using Fisher's linear discriminant.
46. Gaynanova, I. R package version 1.1. 2014. MGSDA: Multi-Group Sparse Discriminant Analysis.
47. Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, Goodacre R. A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding. *Anal Chim Acta*. 2015; 879:10–23. [PubMed: 26002472]
48. Chun H, Keles S. Sparse partial least squares for simultaneous dimension reduction and variable selection. *J R Stat Soc Ser B*. 2010; 72:3–25.
49. Chung D, Keles S. Sparse partial least squares classification for high dimensional data. *Stat Appl Genet Mol Biol*. 2010; 9:Article 17.
50. Habash-Bseiso DE, Yale SH, Glurich I, Goldberg JW. Serologic Testing in Connective Tissue Diseases. *Clin Med Res*. 2005; 3:190–193. [PubMed: 16160075]
51. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw*. 2008; 28:1–26.

HIGHLIGHTS

- New approaches for the diagnostic analysis of thermograms were evaluated
- Classification performance was assessed using a large dataset of lupus and controls
- Thermogram feature metrics and principal components performed modestly
- Classification performance was higher for modern classification algorithms
- Uncovering biological drivers of thermogram changes can enhance diagnostic analysis

Figure 1. Median thermogram and principal component data for lupus and control subjects
 (Left panel) Solid lines represent median thermogram values for lupus and control subjects at each temperature. Confidence bands represent 10th and 90th percentiles for each group of subjects.
 (Right panel) Loadings of the first four principal components for the thermogram data (combined lupus and control samples). The first PC contrasts differences between the heights of the first and third peaks, the second PC is associated with the total area under the thermogram curve, and the third and fourth PCs seemingly involve contrasting thermogram values in the 60–62°C, 69–71°C, 74–76°C, and 80–82°C ranges.

Figure 2. Scatter plot matrix for thermogram peak metrics

Scatter plots are constructed for each pairwise combination of the excess specific heat capacity (J/K.g) for the three prominent peaks in the thermogram data. Points are color-coded according to lupus / control status.

Figure 3. Scatter plot matrix for selected thermogram metrics

Scatter plots are constructed for each pairwise combination of the temperature of the peak maximum ($T_{\max}$), first moment temperature (T_{FM}), and ratio of C_p^{ex} at Peak 1 to C_p^{ex} at Peak 3. Points are color-coded according to lupus / control status.

Figure 4. ROC curves and area under the ROC curve (AUC) values based on the first six principal components of the lupus thermogram data.

Figure 5. ROC curves and area under the ROC curve (AUC) values for six of the calculated summary metrics of the lupus thermogram data.

Figure 6. Accuracy of the six evaluated classification methods for the lupus thermogram data
Box plots represent values from 100 test data sets created by splitting the data randomly into training (two thirds) and testing (one third) sets.

Figure 7. Solution vectors for the six classification methods applied to the lupus thermogram data

In each case the blue line represents the median coefficient value for each thermogram temperature across the 100 training data sets, while the green shaded region represents 10th and 90th percentiles from the 100 training data sets. Training data sets were created by randomly splitting the data into training (two thirds) and testing (one third) sets.